

Remarks

Claims 1-7 and 9-10 are pending in this application. Claims 1-2 and 10 stand rejected under 35 U.S.C. 102(b) as being anticipated by Huang et al. (U.S. Patent No. 5,913,193). Claims 3-5 and 9 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Huang et al. in view of Trader et al. (U.S. Patent No. 5,832,432). Claims 6 and 7 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Huang et al. in view of Holm et al. (U.S. Patent No. 5,850,629).

Claim 1 has been amended to further recite that, in the digital voice library, multiple voice recordings correspond to a single speech item representing various inflections of that single speech item. Further, claim 1 has been amended to recite establishing multiple voice recordings in the digital voice library that correspond to a single inflection of a single speech item, for a plurality of inflections of a plurality of speech items, that represent various ligatures for the single inflection of the single speech item with adjacent speech items.

Accordingly, as amended, claim 1 recites a method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules. The digital voice library includes a plurality of speech items and a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording. Multiple voice recordings corresponding to a single speech item represent various inflections of that single speech item. The method comprises establishing multiple voice recordings in the digital voice library that correspond to a single inflection of a single speech item, for a plurality of inflections of a plurality of speech items, that represent various ligatures for the single inflection of the single speech item with adjacent speech items.

The method further comprises receiving text data, and expanding the text data to form a sequence of text and pseudo words. The method further comprises converting the sequence of text and pseudo words into a sequence of speech items in accordance with the digital voice library. At least one speech item in the sequence of speech items corresponds to

multiple voice recordings. The method further comprises converting the sequence of speech items into a sequence of voice recordings in accordance with the set of playback rules. Selecting a voice recording where multiple voice recordings are available for a speech item is based on context around the speech item in the text data. The voice data is generated based on the sequence of voice recordings by concatenating adjacent recordings in the sequence of voice recordings.

As further recited in claim 1, the plurality of speech items includes a plurality of phrases. Converting the sequence of text and pseudo words further includes parsing the sequence of text and pseudo words to determine any phrases.

With regard to claim 1, Huang fails to describe or suggest the claimed subject matter. Huang fails to describe multiple voice recordings in the digital voice library that correspond to a single inflection of a single speech item, for a plurality of inflections of a plurality of speech items, that represent various ligatures for the single inflection of the single speech item with adjacent speech items, in combination with the further claim limitations.

Huang does describe a method and system of run time acoustic unit selection for speech synthesis. Huang discusses acoustic unit selection for speech synthesis and attempts to minimize the spectral distortion between boundaries of adjacent instances. Huang discusses a training phase where the highest probability instances representing diverse phonetic contexts are chosen. Col. 2, lines 65-67. Huang further describes that the naturalness of synthesized speech is improved by providing multiple instances of an acoustic unit, such as a diphone. Col. 9, lines 58-60.

Huang only describes the creation of a robust library with multiple instances of an acoustical unit but fails to suggest the specific techniques set forth by independent claim 1 involving inflections and ligatures. Claims 2 and 10 depend from claim 1 and are also believed to be patentable. Note that dependent claim 10 depends from claim 9, which depends on claim 1.

With regard to Trader, relied upon by the Examiner as a secondary reference, Trader describes a method for converting a text classified ad to a natural sounding audio ad. Ads from a print media ad database are downloaded to a host processor that identifies relevant ads and converts the abbreviated text of the print ad to an expanded version of the ad. The words and phrases of the expanded version are parsed and stored in appropriate fields of a relational database. A sequence play list of audio file numbers corresponding to the words and phrases contained in a database record is created along with glue words and phrases that are added to produce a more natural sounding audio. Trader fails to describe or suggest, as specifically recited in claim 1, establishing multiple voice recordings in a digital voice library that correspond to a single inflection of a single speech item, for a plurality of inflections of a plurality of speech items, that represent various ligatures for the single inflection of the single speech item with adjacent speech items.

As noted above, Huang has significant shortcomings. As well, Trader has these shortcomings and does not overcome the deficiencies of Huang. With regard to Holm, Holm does not overcome the shortcomings noted above with respect to Huang and Trader.

Accordingly, claims 1-7 and 9-10 are believed to be in condition for allowance and such action is respectfully requested.

Respectfully submitted,
ELIOT M. CASE ET AL.

By /Jeremy J. Curcuri/
Jeremy J. Curcuri
Reg. No. 42,454
Attorney for Applicants

Date: August 30, 2007

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351